$C_d=(D-1)C/2. + D/2. + 0.5; D=2^n$

where:

C is the component value in abstract terms from -1.0 to 1.0

n takes the value 8, 10, or 12, corresponding to the number of bits to be represented.

and C_d is the resulting digital code value.

A signal value of -1.0 results in 0, and a signal value of +1.0 results in 255, 1023, or 4095, corresponding to 8, 10, and 12 bits respectively.

8 Digital timing

8.1 Timing is conveyed using a universal header mechanism. No reserved codes or timing information is contained within the raster.

9 Ancillary data

9.1 No ancillary data is contained within the raster. Ancillary data is conveyed using a universal header mechanism.

Appendix H

Comments To ACATS Meeting on Scanning Formats, 13 July 1995

To: ACATS meeting on Scanning Formats / Compression

13/14 July 1995

From: Gary Demos, Digital Advanced Television Consultant, Apple ATG

Subject: Problems and issues with scanning formats under discussion

The Main Issue: Computer Compatibility

• This issue is being ignored in the current discussions

* This issue is critical

• Without computer interoperability:
  — A separate, "computer-compatible" DTV/ATV system is likely to evolve
  — This different DTV/ATV will be incompatible with the first DTV/ATV
  — Customer confusion will result
  — People will attempt to view text and graphics on screens which are unsuitable, resulting in eyestrain, headaches, even nausea
  — The DTV/ATV system will need to be replaced in very few years (5?)
  — Developers of DTV/ATV programs will need to create multiple variants

• With computer compatibility:
  — One system will be accepted by all
  — Cusdtomers will find that all DTV/ATV devices do what they expect
  — DTV/ATV screens can be used for NII/GII applications (e.g. education)
  — Programs can be developed once for DTV/ATV and computer screens

Encouraging Lines Of Discussion

• De-interlacer to clean up signal at source
• Progressive scan formats at resolutions near NTSC
• Use of progressive scan formats as a source for interlaced NTSC receivers (via set top box line-pair averagers)
• MPEG-2 is acknowledged as very flexible

Discouraging Lines Of Discussion

• Introduction of new interlaced formats
• Introduction of new non-square pixel formats
• Proposed formats don't form a family
• Widescreen being considered at 16:9 aspect ratio
• Formats as point solutions, not forming a system
• Formats for everyone who wants one

• Still talking about the need for interim systems (we are deploying new digital television systems, not interim systems)

Absent Lines of Discussion

• No discussion of overscan
• No discussion of overlay planes
• No discussion of high resolution 1.33:1 aspect ratio, or other aspect ratios (e.g. 2.0:1)
• No discussion of the serious problems associated with 30 and 60 Hz, since computer compatibility requires display rates > 70 Hz
• No discussion about the requirements of uniting computer displays and the new digital video system formats (do people not consider this a requirement? how could this be? Isn't this needed by the N.I.I. and G.I.I.?)
• No discussion of how to prevent proposals for computer-incompatible interim systems from becoming the permanent new systems
• No discussion of the cost increases at the display and quality loss associated with the ATV proposed formats, as well as new formats under discussion
• No discussion of how toake the formats into a coherent and optimal system (e.g. hierarchical/layered compression)

What's Right With The Current Proposals

• MPEG-2 tests well
• Fully digital transmission
• Square pixel formats (some)
• Non-interlaced (progressive scan) formats (some)
• 24 frame per second formats (for movies)

What's Wrong With The Current Proposals

• 60 Hz and 30 Hz, when >70 Hz is needed
• Interlaced formats
• Non-square pixel formats
• 16:9 aspect ratio for widescreen
• Colorimetry limited to TV phosphor colors

What's Missing In The Current Proposals

• Provision for overlay planes
• No high resolution 1.33:1 formats
• Need a prohibition on overscan or precise rules for visible area
• Prohibition on image cropping (pan-and-scan) except at the receiver/display
• Device-indepdent colorimetry
• Definition of clean-signal formats
(reversible transformations must exist for all processes)
• Prohibition on the use of 3-2 pulldown
• A digital interface specification is needed (we recommend P1394/FireWire)
• Error-free data and code transmission is needed

Appendix I

DemoGraFX SMPTE Presentation, 3 Febrary 1996, Describing DemoGraFX ATV System Technical Principles

Temporal and Resolution Layering In Advanced Television

By Gary Demos
DemoGraFX
Santa Monica, CA

Abstract

Current proposals for Advanced Television for the United States are based upon the premise that temporal and resolution layering are inefficient. These proposals therefore only provide a menu of individual formats from which to select, but each format only encodes and decodes a single resolution and frame rate. In addition, it is being suggested by some people that interlace is required, due to their claimed need to have one thousand lines at high frame rates, but based upon the notion that such images cannot be compressed within the available 18mbits/second.

This paper discusses an approach to image compression which demonstrably achieves thousand line image compression at high frame rates with high quality. It also achieves *both* temporal and spatial scalability at this resolution at high frame rates within the available 18mbits/second. This technique efficiently encodes 2 MegaPixel images at 72 frames per second, achieving over twice the compression ratio being proposed by ACATS for advanced television. Further, this proposed technique is more robust than the current unlayered ACATS format proposal for advanced television, since all of the bits may be allocated to the lower resolution base layer when stressful image material is encountered.

Thus, a number of key technical attributes are provided by this proposal, allowing substantial improvement over the ACATS proposal. These improvements include: the replacement of numerous resolutions and frame rates with a single layered resolution and frame rate; no need for interlace in order to achieve a thousand lines of two million pixels at high frame rates; and compatibility with computer displays through the use of 72 frames per second.

Introduction

It would be highly desirable if the digital advanced television standard that the United States adopts to replace our existing NTSC television were to be both flexible and capable. The current proposal under consideration does not provide a crucial capability of compatibility with computer displays. The current proposal also contains a number of specific formats, which are not integrally related to each other. It would be much more desirable if a single digital signal format were to be adopted, containing within it all of the desired standard and high definition resolutions. Temporal (frame rate) and spatial (resolution) scalability would provide such a construction. Unfortunately, the temporal and spatial scalability features specified within MPEG-2 are not sufficiently efficient to operate within the needs of advanced television for the United States. This discussion, however, presents mechanisms which can provide both spatial and temporal scalability at 2 Million pixels, and high frame rates (72 Hz), within the data rate available within a 6 MHz television channel (19 mbps).

As of this writing, ACATS is proposing that the United States adopt digital standard-definition and advanced television formats at rates of 24 Hz, 30 Hz, 60 Hz, and 60 Hz interlaced. It is apparent that these rates are intended to continue the existing television display rate of 60 Hz (or 59.94 Hz). It is also apparent that "3-2 pulldown" is intended for display on 60 Hz displays when presenting movies, which have a temporal rate of 24 frames per second.

These proposed image motion rates are based upon historical rates which date back to the early part of this century. If a "clean-slate" were to be made, it is unlikely that these rates would be chosen. In the computer industry, where displays could utilize any rate over the last decade, rates in the 70 to 80 Hz range have proven optimal, with 75 Hz being the most common rate. Given our historical legacy of high resolution motion pictures at 24 frames per second, the rate of 72 Hz is also implied for consideration.

Unfortunately, the proposed rates of 30 and 60 Hz lack useful interoperability with 72 or 75 Hz, resulting in degraded temporal performance.

Goals Of A Temporal Rate Family

The following goals are therefore in need of consideration in specifying the temporal characteristics of our future digital television systems:

- Optimal presentation of our high resolution legacy of 24 frame-per-second movies

- Smooth motion capture for rapidly moving image types such as sports

- Smooth motion presentation of sports on existing analog NTSC displays, as well as computer-compatible displays operating at 72 or 75 Hz

- Reasonable but more efficient motion capture of less-rapidly-moving images such as news and live drama

- Reasonable presentation of all new digital types of images through a converter box onto existing NTSC displays

- High quality presentation of all new digital types of images on computer-compatible displays

- If 60 Hz digital standard or high resolution displays come into the market, reasonable or high quality presentation on these displays may be required as well.

Since the 60 Hz and 72/75 Hz displays are fundamentally incompatible at any rate other than the movie rate of 24 Hz, the best situation would be if either 72/75 or 60 were eliminated as a display rate. Since 72 or 75 Hz is a required rate for N.I.I. and computer applications, the elimination of the 60 Hz rate as being fundamentally obsolete would be the most future-looking. However,

there are many political forces within the broadcasting and television equipment industries who are insisting that we deploy a new digital television infrastructure based around 60 Hz (and 30 Hz). This has lead to much heated debate between the television, broadcast, and computer industries. Further, the insistence by some members of the broadcast and television industries on *interlaced* 60 Hz formats further widens the gap with computer display requirements.

Interlace

Since non-interlaced display is required for computer-like applications of digital television systems, a de-interlacer is required when interlaced signals are displayed. There is substantial debate about the cost and quality of de-interlacers, since they would be needed in every such receiving device. Frame rate conversion, in addition to de-interlacing, further impacts cost and quality. Note, for example, that NTSC to-from PAL converters continue to be very costly and yet conversion performance is not dependable on many common types of scenes.

In this paper, the issue of interlace is not considered in any detail, as it is a complex and problematic subject. In order to attempt to address the problems and issue of temporal rate, a digital television world without interlace is assumed. It is recognized that this is a dangerous assumption, given the political forces who are insisting on deploying new digital interlaced formats. However the problems of interlace are very difficult, yielding temporal complexity which can only be dealt with adequately by a much more lengthy discussion. Therefore, this discussion proceeds on the basis of evaluating temporal issue in the absence of interlace.

Selecting Optimal Rates

It is certainly true that optimal presentation on a 72 or 75 Hz display will occur if a camera or simulated image is created at 72 or 75 Hz, respectively. Similarly, optimal motion fidelity on a 60 Hz display will result from a 60 Hz camera or simulated image. Use of 72 Hz or 75 Hz with 60 Hz results in a 12 Hz or 15 Hz beat frequency, respectively This beat can be removed through motion analysis, but motion analysis is expensive and inexact, often leading to visible artifacts and temporal aliasing. In the absence of motion analysis, the beat frequency dominates the perceived display rate, making the 12 or 15 Hz beat appear to provide less accurate motion than even 24 Hz. Thus, 24 Hz forms the natural temporal common denominator between 60 and 72 Hz. Although 75 Hz has a slightly higher 15 Hz beat with 60 Hz, its motion is still not as smooth as 24 Hz, and there is no relationship between 75 Hz and 24 Hz unless the 24 Hz rate is increased to 25 Hz. In European 50 Hz countries, movies are often played 4% fast at 25 Hz, and this could be done to make film presentable on 75 Hz displays.

The question remains as to whether there is a higher temporal rate, yielding smoother motion on both 60 Hz and 72 or 75 Hz displays. In the absence of motion analysis at each receiving device, 60 Hz motion on 72 or 75 Hz displays, and 75 or 72 Hz motion on 60 Hz displays will be less smooth than 24 Hz images. Thus, neither 72/75 nor 60 Hz motion is suitable for reaching a heterogeneous display population containing both 72 or 75 Hz and 60 Hz displays.

3-2 Pulldown

There is further complication due to the use of 3-2 pulldown combined with video effects during the telecine process. By some estimates, more than half of all film on video has substantial portions where adjustments have been made at the 59.94 Hz video field rate to the 24 frame-per-second film. Such adjustments include pan-and-scan, color correction, and title scrolling. Further, many films are time-adjusted by dropping frames or clipping the starts and ends of scenes to fit within a given broadcast scheduled These operations can make the 3-2 pulldown impossible to reverse, since there is both 59.94 and 24 Hz motion. This can make the film very difficult to compress using MPEG. Fortunately, this problem is limited to existing NTSC-resolution material, since there is not yet any significant library of higher resolution digital film using 3-2 pulldown.

Motion Blur

In order to further explore the issue of finding a common temporal rate higher than 24 Hz, it is useful to mention motion blur in the capture of moving images. Camera sensors and motion picture film is open to sensing a moving image for a portion of the duration of each frame. On motion picture cameras and many video cameras, the duration of this exposure is adjustable. Film cameras require a period of time to advance the film, and are usually limited to approximately 210 out of 360 degrees, or a 58% duty cycle. On video cameras having CCD sensors, some portion of the frame time is often required to "read" the image from the sensor. This can vary from 10% to 50% of the frame time. In some sensors, an electronic shutter must be used to blank the light during this readout time. Thus, the "duty cycle" of CCD sensors usually varies from 50 to 90%, and is adjustable in some cameras. The light shutter can sometimes be adjusted to further reduce the duty cycle, if desired. However, for both film and video, the most common sensor duty cycle duration is 50%.

With this issue in mind, one can consider the use of only some of the frames from an image sequence captured at 60, 72, or 75 Hz. Utilizing one frame in two, three, four, etc. we have sub rates shown in table 1.

Rate	1/2 Rate	1/3 Rate	1/4 Rate	1/5 Rate	1/6 Rate
75 Hz	37.5	25	18.25	15	12.5
72 Hz	36	24	18	14.4	12
60 Hz	30	20	15	12	10

Table 1
Sub Rates

The rate of 15 Hz is a unifying rate between 60 and 75 Hz. The rate of 12 Hz is a unifying rate between 60 and 72 Hz. However, the desire for a rate above 24 Hz eliminates these rates. 24 Hz is not common, but the use of 3-2 pulldown has come to be accepted by the industry for presentation on 60 Hz displays. The only candidate rates are therefore 30, 36, and 37.5 Hz. Since 30 Hz has a 7.5 Hz beat with 75 Hz, and a 6 Hz beat with 72 Hz, it is not useful as a candidate.

Thus, the rates of 36 and 37.5 Hz become our candidates for smoother motion than 24 Hz material, when presented on 60 and 72 or 75 Hz. These rates are about 50% faster and smoother than 24 Hz. The rate or 37.5 Hz is not suitable for use with either 60 or 72 Hz, so it must be eliminated. This leaves only 36 Hz unless 60 Hz can move 4% to 62.5 Hz. Given the political push behind 60.0 Hz, 62.5 Hz appears unlikely. There are even those who propose the very obsolete 59.94 Hz rate for new television systems.

Thus, we are left with 24, 36, 60, and 72 Hz as candidates for a temporal rate family. 72 and 60 Hz cannot be used for a distribution rate, since motion is less smooth between these two rates than if 24 Hz is used. We therefore examine 36 Hz as a new candidate capture and image distribution rate.

36 Hz

The 3-2 pulldown pattern repeats each frame (or field) three times, then twice, three, two, three, two, etc. This is how 24 frame-per-second film is presented on television at 60 Hz (59.94). When considering 36 Hz, each pattern would be repeated in a 2-1-2 pattern. This can be seen as follows in table 2.

Rate 60 Hz	1	2	3	4	5	6	7	8	9	10
24 Hz	1	1	1	2	2	3	3	3	4	4
36 Hz	1	1	2	3	3	4	4	5	6	6

Table 2

This pattern can also be seen in Figure 1.

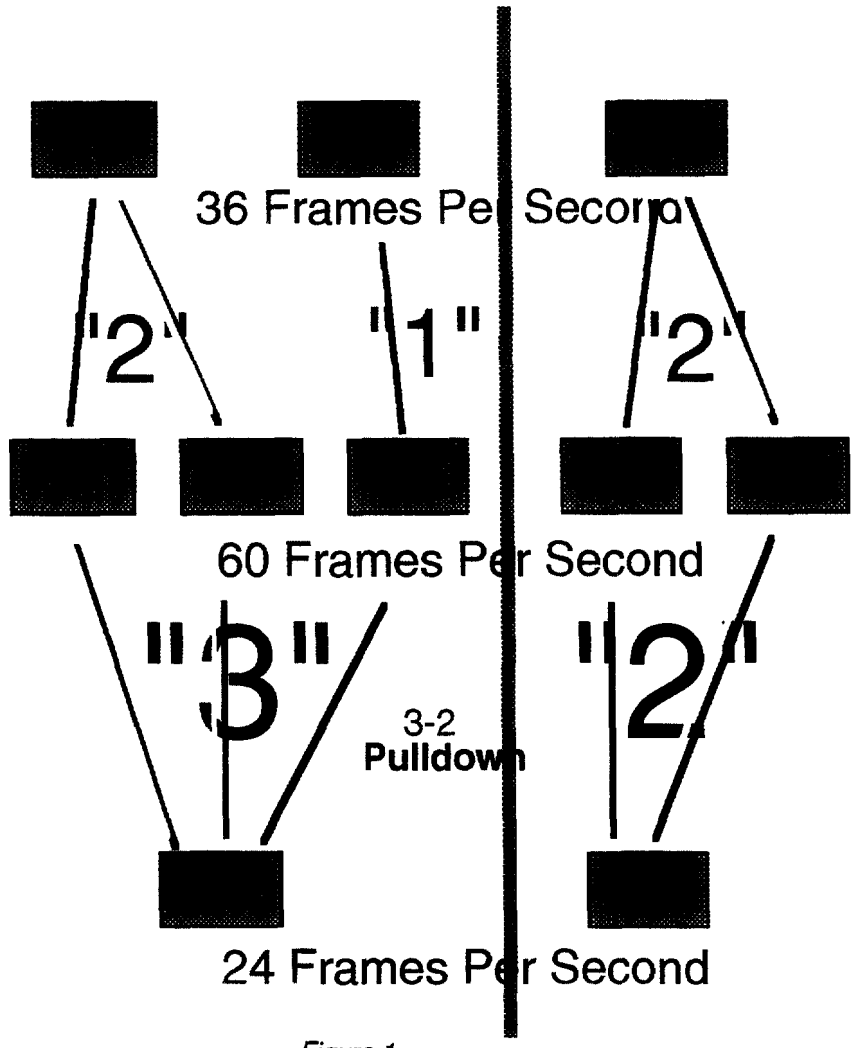

Figure 1
3-2 Pulldown vs 2-1-2 Pulldown

This relationship between 36 Hz and 60 Hz only holds for true 36 Hz material. 60 Hz material can be "stored" in 36 Hz, if it is interlaced, but 36 Hz cannot be reasonable created from 60 Hz without motion analysis and reconstruction. However, in looking for a new rate for motion capture, 36 Hz provides slightly smoother motion on 60 Hz than does 24 Hz, and provides substantially better image motion smoothness on a 72 Hz display. The motion capture rate to 36 Hz thus forms an improvement over 24 Hz on both 60 and 72 Hz displays.

Since 36 Hz cannot be simply extracted from 60 Hz, 60 Hz does not provide a suitable rate for capture. However, we can consider capture at 72 Hz, and utilizing every other frame as the basis of 36 Hz motion. The motion blur at 36 Hz capture will be twice as extensive as the motion blur from every other frame of 72 Hz.

Tests of motion blur appearance of every third frame from 72 Hz show that the staccato strobing at 24 Hz is objectionable. However, utilizing every other frame from 72 Hz at 36 Hz is not objectionable to the eye compared to 36 Hz native capture.

Thus, 36 Hz affords the opportunity to provide very smooth motion on 72 Hz displays by capturing at 72 Hz, while providing better motion on 60 Hz displays than 24 Hz via utilizing 36 Hz alternate frames.

Thus, the temporal rates for capture and distribution shown in Figure 2 appear optimal.

Capture Rate	Distribution Rate	Optimal Display Rate	Acceptable Display
72 Hz	36 Hz + 36 Hz	72 Hz	60 Hz

Figure 2.
Optimal Temporal Rates

It is worth noting that this technique of utilizing alternate frames from a 72 Hz camera to achieve a 36 Hz base rate can profit from an increased motion blur duty cycle. The normal 50% duty cycle at 72 Hz, yielding a 25% duty cycle at 36 Hz has been demonstrated to be acceptable, and to represent a significant improvement over 24 Hz on 60 Hz and 72 Hz displays. However, if the duty cycle could be increased to 75% or perhaps 90%, then the 36 Hz samples would begin to approach the more common 50% duty cycle. It is possible with "backing store" CCD designs to have a short blanking time, yielding a high duty cycle.

MPEG-2 Coding Pattern

When using MPEG compression, it is possible to embed a simple temporal layer between the 36 Hz rate and the 72 Hz rate if the "P-frame" distance is even. Thus, the MPEG coding patterns of PBPBPBPB or PBBBPBBBPBBBP will both afford placing alternate frames in a separate stream containing only the temporal enhancement B frames to take 36 Hz to 72 Hz. Craig Birkmaier suggested this construction for achieving temporal layering within MPEG-2 compression. These coding patterns are shown in Figures 3 and 4.

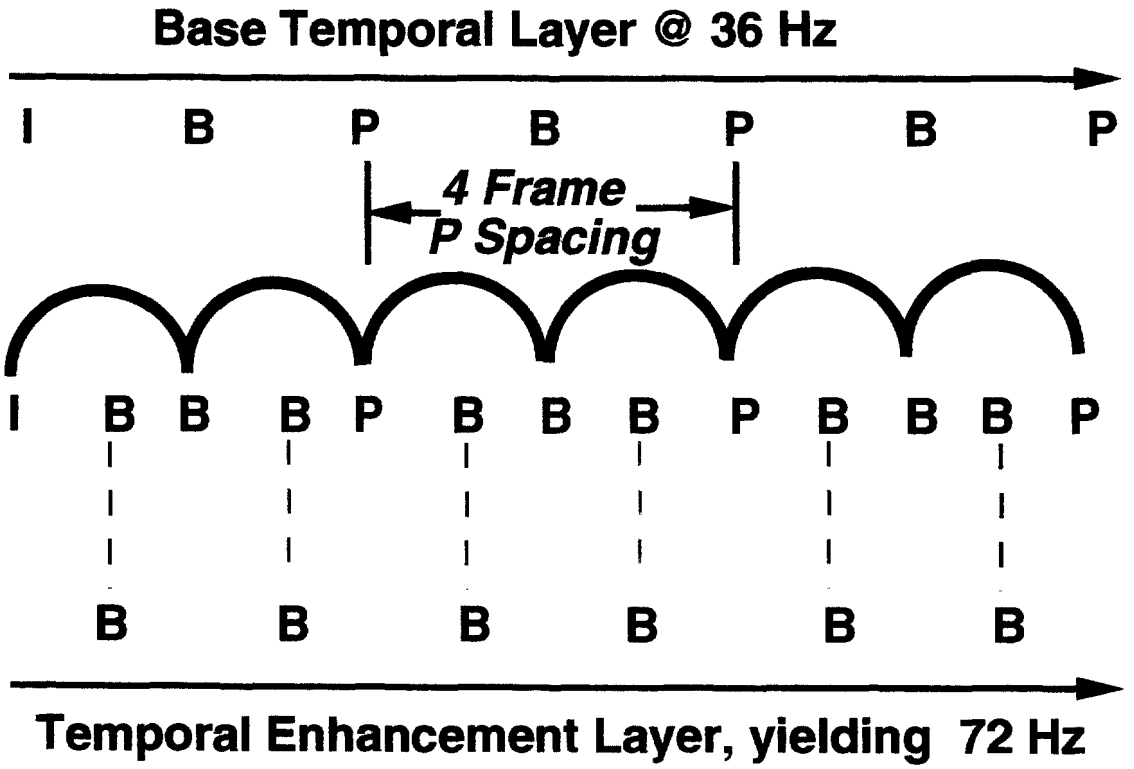

Figure 3
36/72 Hz Temporal Base and Enhancement Structure
4-Frame P Spacing

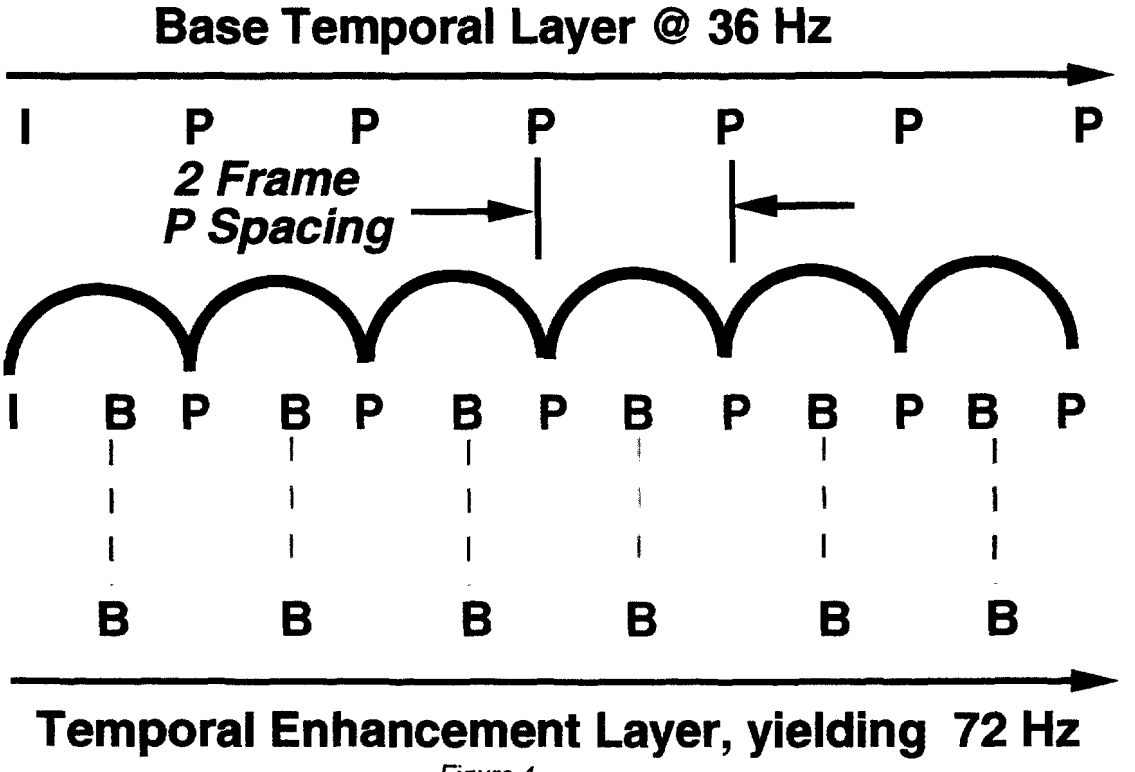

Figure 4
36/72 Hz Temporal Base and Enhancement Structure
2-Frame P Spacing

The construction of Figure 4 has the added advantage that the 36 Hz decoder would only need to decode "P" frames, reducing the required memory bandwidth if 24 Hz movies were also decoded without "B" frames. Experiments with high resolution images have suggested that the 2-Frame P spacing of Figure 4 is optimal for most types of images.

In tests, the construction in Figure 4 appears to offer the optimal temporal structure for supporting both 60 and 72 Hz, while providing excellent results on the modern 72 Hz computer-compatible displays.

This construction allows two digital streams, one for the base layer at 36 Hz, and one for the enhancement layer B frames to achieve 72 Hz. This is illustrated in Figure 5.

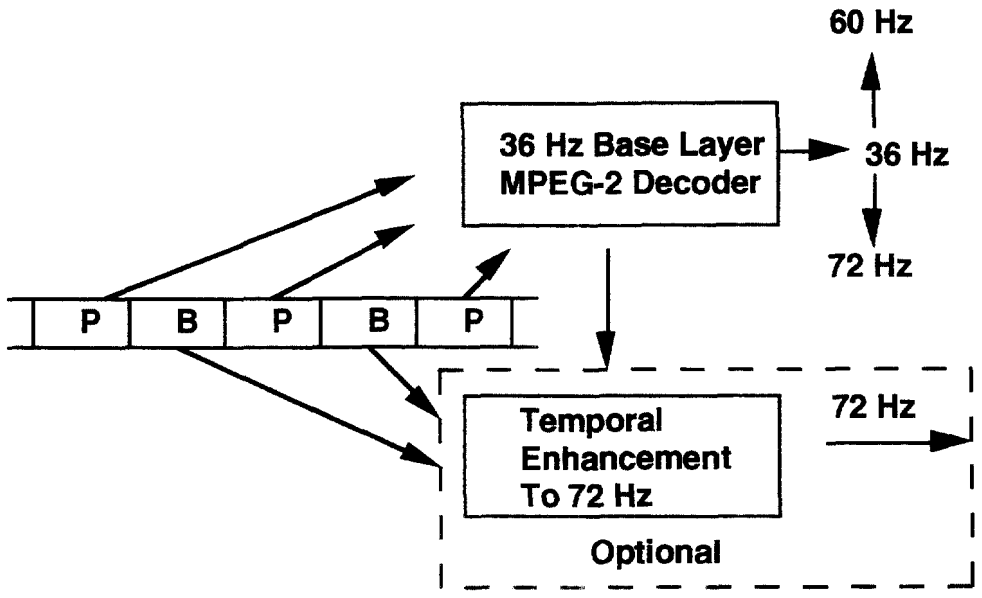

Figure 5.
MPEG -2 Temporal Layer Decoding

Existing 60 Hz Interlaced Material

Most existing 60 Hz interlaced material is video tape for NTSC in analog, D1, or D2 format. There is also a small amount of Japanese HDTV (aka SMPTE 240/260M). There are also cameras which operate in this format. Any such 60 Hz interlaced format can be processed with a "fancy box" whereby the signal is de-interlaced and frame rate converted. This process involves very complex image understanding technology, similar to robot vision. Even with very sophisticated technology, temporal aliasing will result in "misunderstandings" by the algorithm, and occasionally yield artifacts. Note that the typical 50% duty cycle of image capture means that the camera is "not looking" half the time. The "backwards wagon wheels" in movies is an example of temporal aliasing due to this normal practice of temporal undersampling. Such artifacts cannot be removed without human-assisted reconstruction. Thus, there will always be cases which cannot be automatically corrected. When motion analysis results in machine understanding, it will be the equivalent of having robots that can see and understand. This is certainly not likely in the near future. For a very sophisticated fancy box, however, the motion conversion results available in current technology should be reasonable on most material.

The price of a single high definition camera or tape machine would be similar to the cost of such a fancy box. Thus, in a studio having several cameras and tape machines, the cost of such conversion becomes modest. However, performing such processing adequately is beyond the budget of home and office products. Thus, the complex processing to remove interlace and convert the frame rate belongs at the origination studio. This is shown in Figure 12.

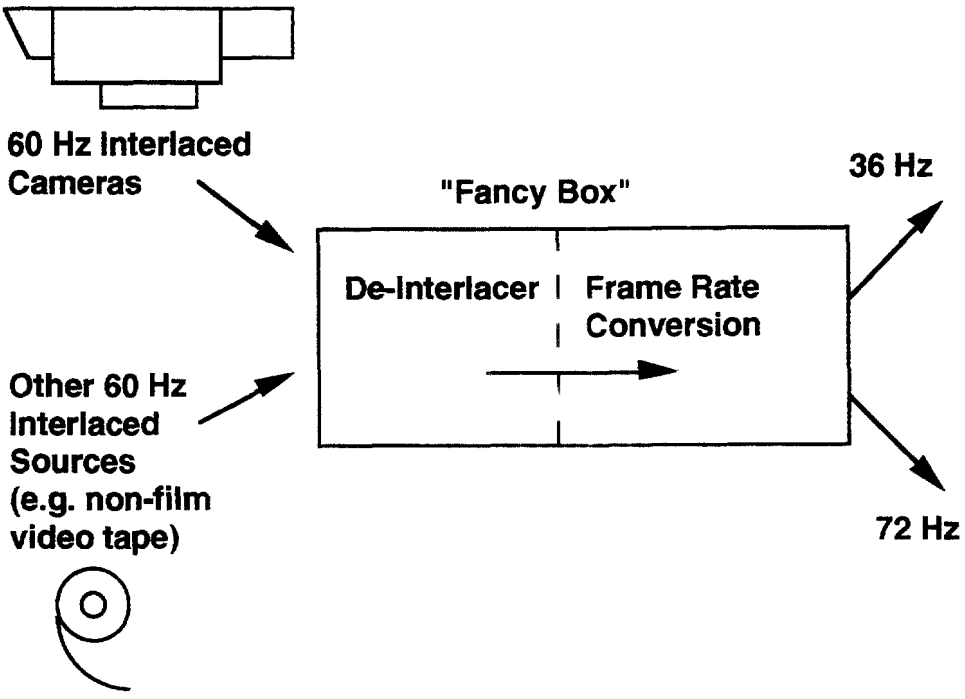

Figure 12
"Fancy Box" Motion Analysis To Yield 36/72 Hz

This process can also be adapted to produce a second temporal enhancement layer on the 36 Hz base layer which would reproduce the original 60 Hz, although de-interlaced. If similar quantization is used for the enhancement B frames, the data rate should be slightly less than the 72 Hz enhancement layer, since there are fewer B frames. However, this use of the data bandwidth should probably be discouraged, since it encourages the use of new 60 Hz non-interlaced receivers, which are not suitable for text and graphics. Thus, it is recommended that only the 72 Hz enhancement layer be present, if any enhancement layer is used.

The same process can also be applied to the conversion of existing PAL 50 Hz material. PAL video tapes are best slowed to 48 Hz prior to this conversion. Live PAL requires processing with the relatively unrelated rates of 50, 36, and 72 Hz. Such units are only affordable at the source of broadcast signals, and are not practical at each receiving device in the home and office.

The vast majority of material of interest to the United States is low resolution NTSC. At present, most NTSC signals are viewed with substantial impairment on most home televisions. Further, viewers have come to accept the temporal impairments inherent in the use of 3-2 pulldown to present film on television. Nearly all prime-time television is made on film at 24 frames per second. Thus, only sports, news, and other video-original shows need be processed. The artifacts and losses associated with converting these shows to a 36/72 Hz format are likely to be offset by the improvements associated with high-quality de-interlacing of the signal.

Note that the motion blur inherent in the 60 Hz (or 59.94) fields should be very similar to the motion blur in 72 Hz frames. Thus, the process should appear similar to 72 Hz origination in terms of motion blur.

Thus, few viewers will notice the difference, except possibly as a slight improvement, when their interlaced 60 Hz NTSC viewing of old material is processed into 36 Hz. However, those who buy new 72 Hz digital non-interlaced televisions will notice a small improvement when viewing NTSC, and a major improvement when viewing new material captured or originated at 72 Hz. Even base-level decoding of 36 Hz presented on 72 Hz displays will look as good as high quality digital NTSC, replacing interlace artifacts with a slower frame rate.

Main Level, Main Profile, MPEG-2 Decoders

A number of companies are building MPEG-2 decoding chips which operate at around 11 MPixels/second. MPEG-2 has defined some "profiles" for resolutions and frame rates. Although these profiles are strongly biased toward computer-incompatible format parameters such as 60 Hz, non-square pixels, and interlace, many chip manufacturers appear to be developing decoder chips which operate at the "main profile, main level". This profile is defined to be any horizontal resolution up to 720 pixels, any vertical resolution up to 576 lines at up to 25 Hz, and any frame rate of up to 480 lines at up to 30 Hz. A wide range of data rates from approximately 1.5 Mbits/second to about 10 Mbits/second is also specified. However, from a chip point of view, the main issue is the rate at which pixels are decoded. The main-level, main-profile pixel rate is about 10.5 MPixels/second.

Although there is variation among chip manufacturers, most MPEG-2 decoder chips will operate at up to 13 MPixels/second, given quick support memory. Some decoder chips will go as fast as 20 MPixels/second or more. Given that CPU chips tend to gain 50% improvement or more each year at a given cost, one can expect some near-term flexibility in the pixel rate of MPEG-2 decoder chips.

Table 3 illustrates some desirable resolutions and frame rates, and their corresponding pixel rates.

Resolution X	Y	Frame Rate (Hz)	Pixel Rate (MPixels/s)
640	480	36	11.1
720	486	36	12.6
720	486	30 *(for comparison)*	10.5
704	480	36	12.2
704	480	30 *(for comparison)*	10.1
680	512	36	12.5
1024	512	24	12.6

Table 3
Base Layer Formats Near Main-Level/Main-Profile MPEG-2

Notice that all of these formats can be utilized with MPEG-2 decoder chips which can provide 12.6 MPixels/second. The very desirable 640 x 480 at 36 Hz format can be achieved by nearly all chips, since its rate is 11.1 MPixels/second. A widescreen 1024 x 512 image can be squeezed into 680 x 512 using a 1.5:1 squeeze, and can be supported at 36 Hz if 12.5 MPixels/second can be handled. The highly desirable square pixel widescreen template of 1024 x 512 can achieve 36 Hz if MPEG-2 decoder chips can achieve 18.9 MPixels/second. This becomes more feasible if 24 Hz and 36 Hz material is coded only with P frames, such that B frames are only required in the 72 Hz temporal enhancement layer decoders. Decoders which use only P frames require less memory and memory bandwidth, making the goal of 19 MPixels/second more accessible.

The 1024 x 512 resolution template would most often be used with 2.35:1 and 1.85:1 movies at 24 frames per second. These only require 11.8 MPixels/second, which should fit within the limits of most existing main level-main profile decoders. These formats are shown together in a "master template" for a base layer at 24 or 36 Hz in Figure 6.

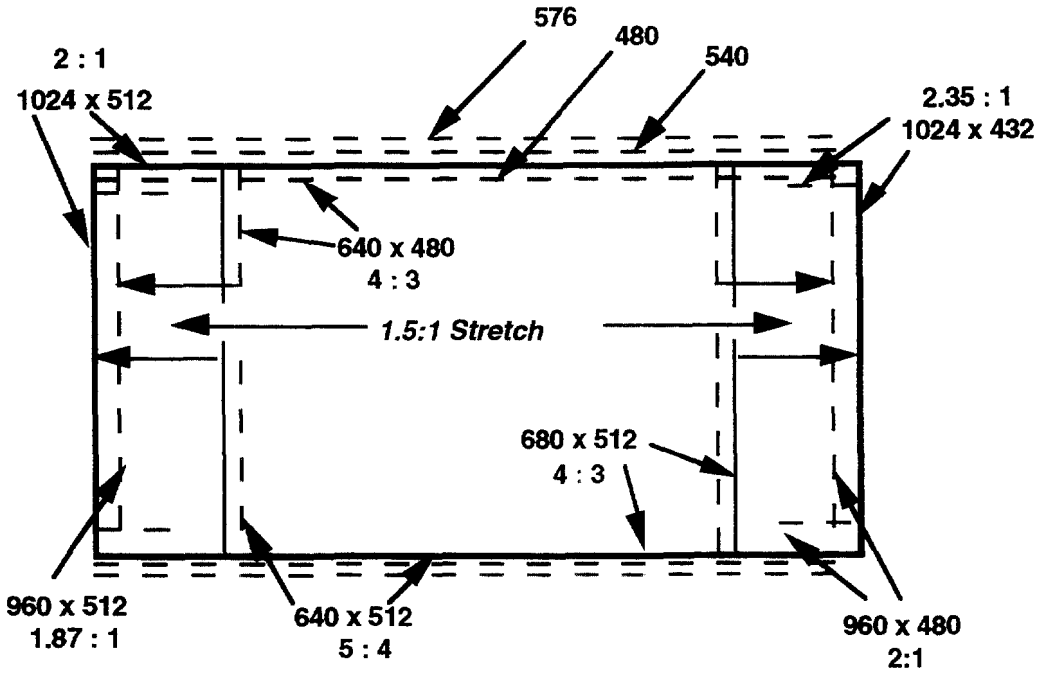

Figure 6
Base Layer Master Resolution Template For 24 and 36 Hz

The temporal enhancement layer of "B" frames to present 72 Hz can be decoded using a chip with double the pixel rates specified above, or by using a second chip in parallel with additional access to the decoder memory. The merging of the enhancement and base layer stream to insert the alternate "B" frames can be done invisibly to the decoder chip using the MPEG-2 transport layer. The transport packets for two PID's can be recognized as containing the base layer and enhancement layer, and their stream contents can both be simply passed on to a double-rate capable decoder chip, (or to an appropriately configured normal rate pair of decoders). It is also possible to use the "data partitioning" feature in the MPEG-2 data stream instead of the transport layer from MPEG-2 systems. The data partitioning feature allows the B frames to be marked as belonging to a different class within the MPEG-2 compressed data stream, and can therefore be flagged to be ignored by 36-Hz decoders which only support the temporal base layer rate.

It should be noted that temporal scalability, as defined by MPEG-2 video compression, is not as optimal as the simple B frame partitioning proposed here. The MPEG-2 temporal scalability is only forward referenced from a previous P or B frame, and thus lacks the efficiency available in the B frame encoding proposed here, which can is both forward and backward referenced.

Thus, the simple use of B frames as the temporal enhancement layer provides a simpler and more efficient temporal scalability than does the temporal scalability defined within MPEG-2. Also, this use of B frames as the mechanism for temporal scalability is fully compliant with MPEG-2. The two methods of identifying these B frames as an enhancement layer, via data partitioning or alternate PID's for the B frames, are also fully compliant.

Resolution Scalability

It is possible to enhance the base resolution template using hierarchical resolution scalability utilizing MPEG-2 to achieve higher resolutions built upon this base layer. Use of enhancement can achieve resolutions at 3/2 and double the base layer. The double resolution can be built in two steps, by using 3/2 then 4/3, or it can be a single factor-of-two step. This is shown in Figure 7.

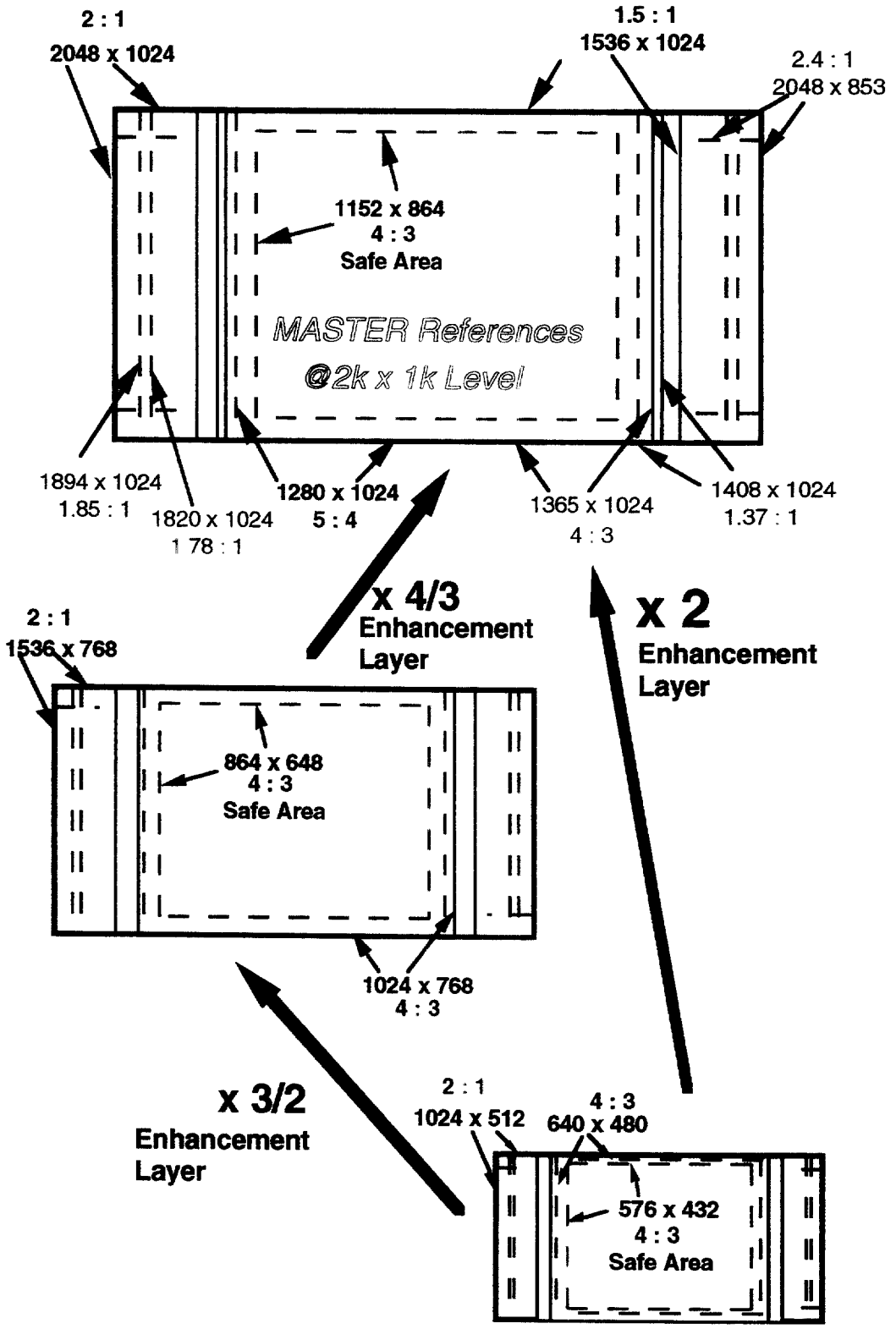

Figure 7
Resolution Enhancement From The Base Layer

The process of resolution enhancement can be achieved by utilizing MPEG-2 compression on the enhancement layers by treating them as independent MPEG-2 streams. This technique differs from the "spatial scalability" defined with MPEG-2. Spatial scalability with MPEG-2 has proven to be highly inefficient, and is best not used. However, MPEG-2 contains all of the tools to construct an effective layered resolution to provide spatial scalability.

The layered resolution encoding process is shown in Figure 8.

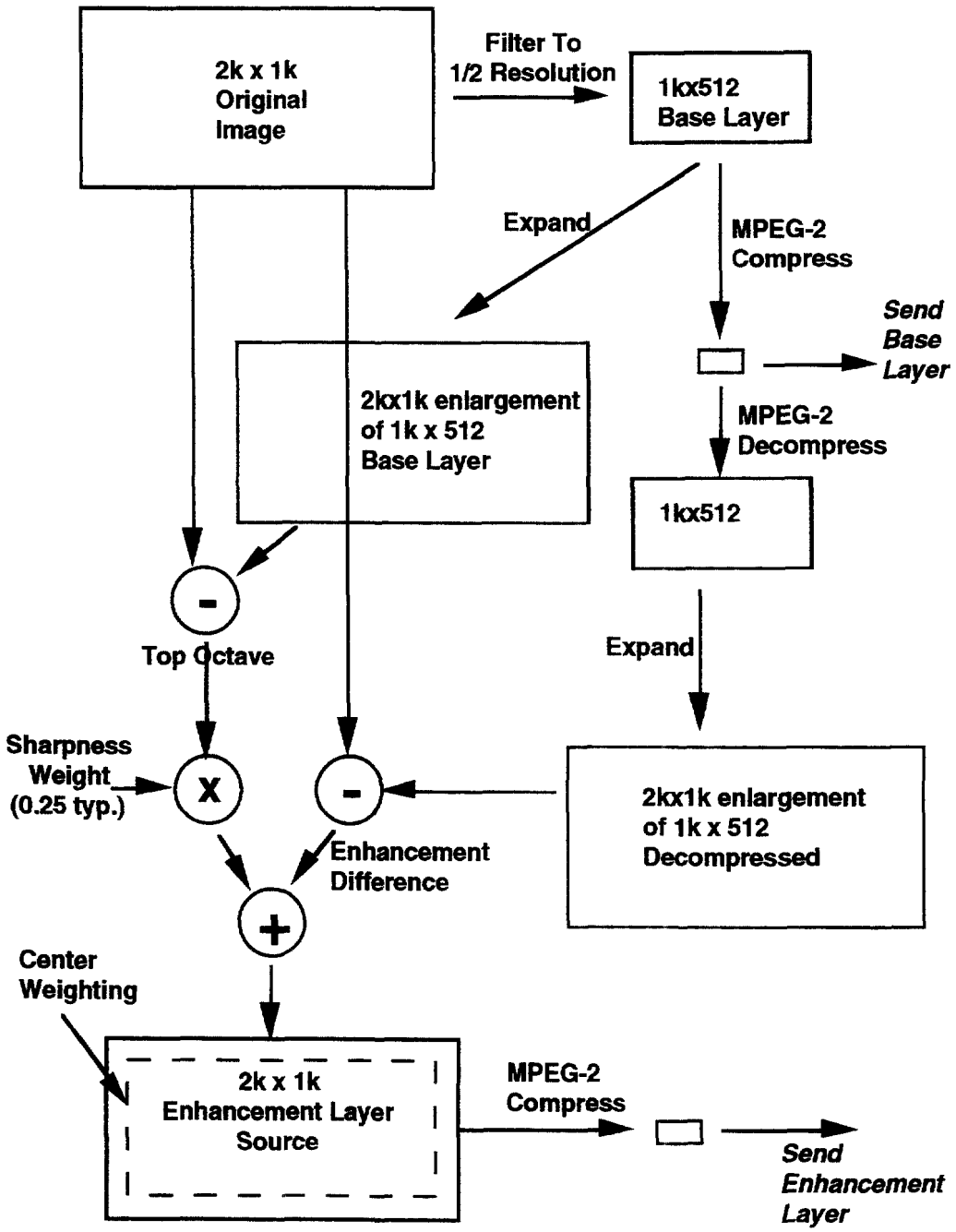

Figure 8
Layered Resolution Encoding Process

The decoding process is shown in Figure 9.

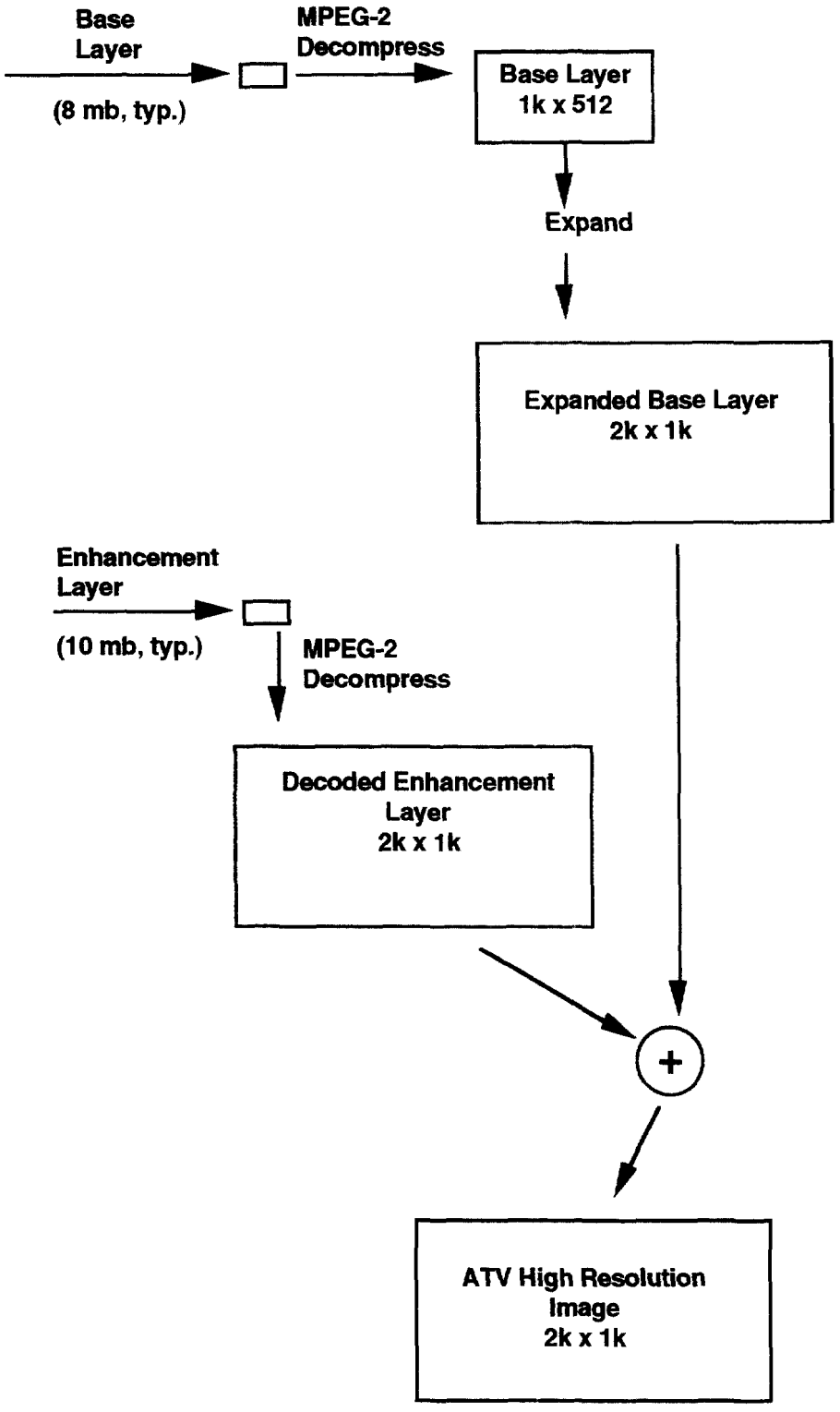

Figure 9
Layered Resolution Decoding Process

As with MPEG-2, the enhancement layer is created by expanding the decoded lower layer, taking the difference, and then compressing. The compressed spatial resolution enhancement layer may be optionally added to the base layer after decoding to create a higher resolution image in the decoder.

However, this layered resolution encoding process differs from MPEG-2 spatial scalability in several ways.

• The enhancement layer *difference picture* is compressed as its own MPEG-2 data stream, with I, B, and P frames. This difference represents the major reason that resolution scalability, as proposed here, is effective, where MPEG-2 spatial scalability is ineffective. The spatial scalability defined within MPEG-2 allows the upper layer to be coded as the difference between the upper layer picture and the expanded base layer, or as a motion compensated MPEG-2 data stream of the *actual picture* or a combination of both. However, neither of these encodings is efficient. The difference from the base layer could be considered as an I-frame of the difference, which is inefficient compared to a motion-compensated difference picture, as proposed here. The upper-layer encoding defined within MPEG-2 is also inefficient, since it is identical to a complete encoding of the upper layer. The motion compensated encoding of the difference picture, as proposed here, is therefore substantially more efficient.

• Since the enhancement layer is an independent MPEG-2 data stream, the MPEG-2 systems transport layer (or another similar mechanism) must be used to multiplex the base and enhancement layers.

• The expansion and resolution reduction filtering can be a gaussian or spline function, which are more optimal than the bilinear interpolation specified in MPEG-2 spatial scalability.

• The image aspect ratio must match between the lower and higher layers in this proposal. In MPEG-2 spatial scalability, extensions to width and/or height are allowed. Such extensions are not allowed in this proposal due to efficiency requirements.

• Due to efficiency requirements, and the extreme amounts of compression used in the enhancement layer, the entire area of the enhancement layer is not coded. Usually, the area excluded from enhancement will be the border area. However, any method of determining the regions having detail which they eye will follow can be utilized to select regions which need detail, and to exclude regions where extra detail is not required. Remember, all of the image has detail to the level of the base layer, so all of the image is present. Only the areas of special interest require the enhancement layer. In the absence of other criteria, the edges of the frame can be excluded from enhancement. The MPEG-2 parameters "lower_layer_prediction_horizontal&vertical offset" parameters used as signed negative integers, combined with the horizontal&vertical_subsampling_factor_m&n values can be used to specify the enhancement layer rectangle's overall size and placement within the expanded base layer.

• A sharpness factor is added here to the enhancement layer to offset the loss of sharpness which occurs during quantization. Care must be taken to utilize this parameter only to restore the clarity and sharpness of the original picture, and not to enhance the image. The picture which adds this sharpness is the "high octave" of resolution between the original high resolution image and the original base layer image. This high octave image will be quite noisy, in addition to containing the sharpness and detail of the high octave of resolution. Adding too much of this image can yield instability in the motion compensated encoding of the enhancement layer. The amount that should be added depends upon the level of the noise in the original image. For noisy images, no sharpness should be added, and it even may be advisable to suppress the noise in the original for the enhancement layer before compressing using noise suppression techniques which preserve detail.

• Temporal and spatial scalability are intermixed by utilizing B frames for temporal enhancement from 36 to 72 Hz in *both* the base and enhancement layers. In this way, four possible levels of decoding performance are possible with two layers of resolution scalability, due to the options available with two levels of temporal scalability.

These differences represent substantial improvements over MPEG-2 spatial and temporal scalability.

These differences are still consistent with MPEG-2 decoder chips, although additional logic is required in the decoder to perform the expansion and addition in the resolution enhancement step. This additional logic is nearly identical to that required by the less effective MPEG-2 spatial scalability.

Optional Non-MPEG-2 Coding Of The Resolution Enhancement Layer

It is also possible to utilize a different compression technique for the resolution enhancement layer than MPEG-2. It is not necessary to utilize the same compression technology for the resolution enhancement layer as for the base layer. For example, motion-compensated block wavelets can be utilized to match and track details with great efficiency when the difference layer is coded. Even if the most efficient position for placement of wavelets jumps around on the screen due to changing amounts of differences, it would not be noticed in the low-amplitude enhancement layer. Further, it is not necessary to cover the entire image, it is only necessary to place the wavelets on details. The wavelets can have their placement guided by detail regions in the image. The placement can also be biased away from the edge.

At the bit rates being described here, where 2 MPixels (2048 x 1024) at 72 frames per second is being coded in 18.5 mbits/second, only a base layer (1024 x 512 at 72fps) and single enhancement layer for resolution have been successfully demonstrated. However, with the anticipated improved efficiencies available from further refinement of enhancement layer coding should allow for multiple enhancement layers. For example, it is conceivable that a base layer at 512 x 256 could be resolution-enhanced by four layers to 1024 x 512, 1536 x 768, and 2048 x 1024. This is possible with existing MPEG-2 coding at the movie frame rate of 24 frames per second. At high frame rates such as 72 frames per second, MPEG-2 does not provide sufficient efficiency in the coding of resolution-enhancement layers to allow this many layers.

The exploration of optimum efficiency and flexibility in encoding resolution-enhancement layers is worth further study.

Graceful Degradation

The techniques described here work well for normal running material at 72 frames per second, with 2k x 1k material. It will also work well on film-based movies, which run at 24 frames per second. At high frame rates, however, when a very noise-like image is coded, or when there are numerous shot cuts within the image stream, the enhancement layer may loose the coherence between frames which is necessary for effective coding. When this happens, it is easily detected, since the buffer-fullness/rate-control mechanism will attempt to set the quantizer to very coarse settings. When this condition is encountered, all of the bits from the enhancement layer should be allocated to the base layer, since the base layer will need as many bits as possible in order to code the stressful material. At between 1/2 and 1/3 MPixel, at 72 frames per second, the resultant pixel rate will be 24 to 36 MPixels/second. This provides 1/2 to 2/3 of a mbit per frame at 18.5 mbits/second, which should be sufficient to code very well, even on stressful material.

However, under more extreme cases, where every frame is very noise-like and/or there are cuts happening every few frames, it is possible to gracefully degrade even further without loss of in the base layer. This can be done by removing the B frames, and thus placing all of the bits in the I and P frames at 36 frames per second. This increases the amount of data available for each frame to between 1 and 1.5 mbits/frame (depending on the resolution of the base layer). This will still yield the fairly good motion rendition rate of 36 frames per second, at the fairly high quality

resolution of the base layer, under these extremely stressful coding conditions. If the base-layer quantizer is still operating at a course level under 18.5 mbits/second at 36 frames per second, then the frame rate can be further reduced to 24, 18, or even 12 frames per second (which would provide between 1.5 and 4 mbits for every frame), which should be able to handle even the most pathological moving image types.

Note that the current proposal for U.S. advanced television (from "ACATS") does not allow for these methods of graceful degradation, and therefore cannot perform as well on stressful material as the system being proposed here.

Mastering Formats

Utilizing a template at or near 2048 x 1024, it is possible to create a single digital moving image master format source for a variety of release formats. As shown here, the 2k x 1k template can efficiently support the common widescreen aspect ratios of 1.85:1 and 2.35:1. It can also accommodate 1.33:1 and other aspect ratios.

Although integers (especially the factor of 2) and simple fractions (3/2 & 4/3) are most efficient step sizes in resolution layering, it is also possible to use arbitrary ratios to achieve any required resolution layering. However, using a 2048 x 1024 template, or something near it, provides not only a high quality digital master format, but also can provide many other convenient resolutions from a factor of two base layer (1kx512), including NTSC, the U.S. television standard.

It is also possible to scan film at higher resolutions such as 4k x 2k, 4k x 3k, or 4k x 4k. Using optional resolution enhancement, these higher resolutions can be created from a central master format resolution near 2k x 1k. Such enhancement layers for film will consist of both image detail, grain, and other sources of noise (such as scanner noise). Because of this noisiness, the use of compression technology in the enhancement layer for these very high resolutions will require alternatives to MPEG-2 types of compression. Fortunately, other compression technologies exist which can be utilized for compressing such noisy signals, while still maintaining the desired detail in the image.

Of course, digital mastering formats should be created in the frame rate of the film if from existing movies at 24 frames per second. The common use of both 3-2 pulldown and interlace are inappropriate for digital film masters.

For new digital electronic shows, it is hoped that the use of 60 Hz interlace will cease in the near future, and be replaced by frame rates which are more compatible with computers, such as 72 Hz, as proposed here. The digital image masters should be made at whatever frame rate the images are captured, whether at 72 Hz, 60 Hz, 36 Hz, 37 5 Hz, 75 Hz, 50 Hz, or other rates.

Combined Spatial and Temporal Enhancement Layers

Inherent in this proposal is the combination of both temporal and resolution enhancement layering.

The temporal enhancement is provided by decoding the "B" frames. The resolution enhancement also contains B frames, so that it also offers two temporal layers.

For 24 frame per second film, the most efficient and lowest cost decoders might use only "P" frames, thereby minimizing both memory and memory bandwidth, as well as simplifying the decoder by eliminating B frame decoding. This is shown in Figure 10.

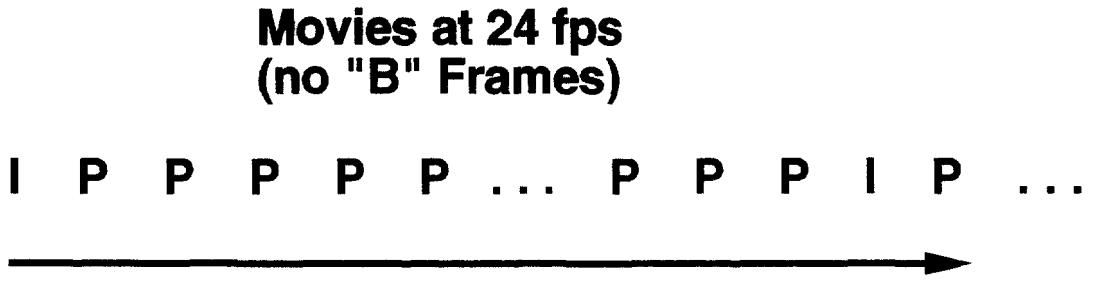

Figure 10
24-Frame-Per-Second Movie Decoding

Thus, decoding of movies at the 24 frame-per-second rate, and decoding of advanced television at 36 frames per second could utilize a decoder without B frame capability. B frames can then be utilized between the P frames to yield the higher temporal layer at 72 Hz as shown in Figure 4.

This layering also applies to the enhanced resolution layer, which can similarly utilize only P (and I) frames for the 24 and 36 frame-per-second rates. The enhancement layer can add the full rate of 72 Hz at high resolution by adding B frame decoding within the enhancement layer.

The combined spatial and temporal scalable options are illustrated in Figure 11.

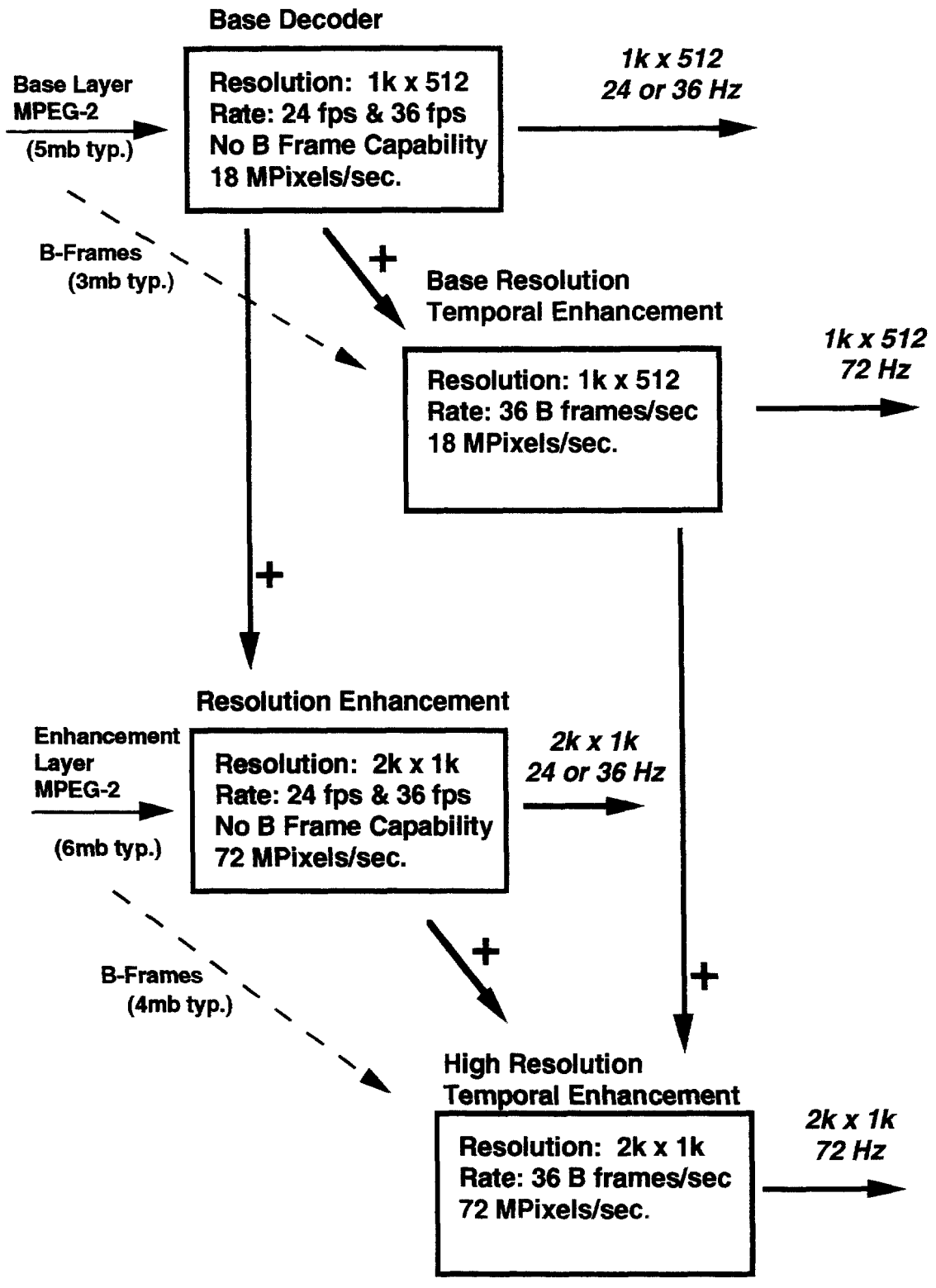

Figure 11
Combined Decoder Temporal and Resolution Scalability Options

This example shows an allocation of the proportions of the 18mbit/second data stream to achieve the spatio-temporal layered Advanced Television

Note that the compression ratio achieved through this proposed scalable encoding mechanism is very high, indicating excellent compression efficiency. These ratios are shown in Table 4 for each of the temporal and scalability options from the example in Figure 11. These ratios are based upon source RGB pixels at 24 bits/pixel. If the 16 bits/pixel of 4:2:2 or the 12 bits/pixel of 4:2:0 encoding are factored in, then the compression ratios would be 3/4 and 1/2 the values shown.

Layer	Resolution	Rate	Data Rate (typ.)	MPixels/s	Comp. Ratio (typ.)
Base	1k x 512	36	5 mb/s	18.9	90
Base Temporal	1k x 512	72	8 mb/s (5+3)	37.7	113
High	2k x 1k	36	11 mb/s (5+6)	75.5	165
High Temporal	2k x 1k	72	18 mb/s (5+3+6+4)	151	201
for comparison:					
CCIR 601	720 x 486	29.97	5 mb/s	10.5	50

Table 4
Compression Ratios

These high compression ratios are enabled by two factors:

1) The high temporal coherence of high-frame-rate 72 Hz images

2) The high spatial coherence of the high resolution 2k x 1k images

3) Application of resolution detail enhancement to the central heart of the image, and not to the borders of the frame.

These factors are exploited in this layered compression technique through taking advantage of the strengths of MPEG-2 encoding. These strengths include the bi-directionally interpolated "B" frames for temporal scalability. MPEG-2 also provides efficient motion representation through the use of motion-vectors in both the base and enhancement layer. Up to some threshold of high noise and rapid image change, MPEG-2 is also efficient at coding the details instead of the noise within the enhancement layer through motion compensation in conjunction with DCT quantization. Above this threshold, the data is best allocated to the base layer. These MPEG-2 mechanisms work together to yield highly efficient and effective coding which is both temporally and spatially scalable.

In comparison to 5 mbit/second encoding of CCIR 601 digital video, the compression ratios can be seen to be much higher. One reason for this is the loss of some coherence due to interlace. Interlace negatively affects both the ability to predict subsequent frames and fields, as well as the correlation between vertically adjacent pixels. Thus, a major portion of the gain in compression efficiency described here is due to the absence of interlace.

These large compression ratios achieved here can be considered from the perspective of the number of bits available to code each MPEG-2 macroblock. A macroblock is a 16 x 16 pixel grouping of four 8 x 8 DCT blocks, together with one motion vector for P frames, and one or two motion vectors for B frames. The bits available per macroblock for each layer are shown in Table 5.

Layer	Data Rate (typ.)	MPixels/s	Average Available Bits/Macroblk
Base	5 mb/s	19	68 bits/macroblock
Base Temporal	8 mb/s (5+3)	38	54 bits/macroblock
High layer	11 mb/s (5+6)	76	37 " overall, 20 enh.
High with enh. border layer	11 mb/s (5+6)	61	46 " overall, 35 enh.
High Temporal layer	18 mb/s (5+3+6+4)	151	30 " overall, 17 enh.
High Tempor. w/border layer	18 mb/s (5+3+6+4)	123	37 " overall, 30 enh.
for comparison:			
CCIR 601	5 mb/s	10.5	122 bits/macroblock

Table 5
Available Bits For Each Macroblock

The available bits to code each macroblock is smaller in the enhancement layer than in the base layer. This is appropriate, since it is desirable for the base layer to have as much quality as possible. The motion vector requires 8 bits or so, leaving 10 to 25 bits for the macroblock type codes and for the DC and AC coefficients for all four 8 x 8 DCT blocks. This leaves room for only a few "strategic" AC coefficients. Thus, statistically, most of the information available for each macroblock must come from the previous frame of the enhancement layer.

It is easily seen why the MPEG-2 spatial scalability is ineffective at these compression ratios, since there is not sufficient data space available to code enough DC and AC coefficients to represent the high octave of detail represented by the enhancement difference image. The high octave is represented primarily in the fifth through eighth horizontal and vertical AC coefficients. These coefficients cannot be reached if there are only a few bits available per DCT block.

The system described here gains its efficiency by utilizing the motion compensated prediction from the previous enhancement difference frame. This is demonstrably effective in providing excellent results in temporal and spatial layered encoding.

In most MPEG-2 encoders, the adaptive quantization level is controlled by the output buffer fullness. At the high compression ratios involved in the enhancement layer, this mechanism may not function optimally. Various techniques can be used to optimize the allocation of data to the most appropriate image regions. The conceptually simplest technique is to perform a pre-pass of encoding over the enhancement layer to gather statistics and to search out details which should be preserved. The results from the pre-pass can be used to set the adaptive quantization to optimize the preservation of detail in the enhancement layer. The settings can also be artificially biased to be non-uniform over the image, such that image detail is biased to allocation in the main screen regions, and away from the macroblocks at the extreme edges of the frame.

Except for leaving an enhancement-layer border at high frame rates, none of these adjustments are required, since existing decoders function well without such improvements. However, these further improvements are available with a small extra effort in the enhancement layer encoder.

Conclusion

The choice of 36 Hz as a new common ground temporal rate appears to be optimal. Demonstrations of the use of this frame rate indicate that it provides significant improvement over 24 Hz for both 60 Hz and 72 Hz display. 36 Hz images can be created by utilizing every other frame from 72 Hz image capture. This allows a base rate of 36 Hz, and a temporal enhancement, using "B" frames, to achieve 72 Hz display.

The "future-looking" rate of 72 Hz is not compromised by this approach, while providing transition for 60 Hz analog NTSC display. It also allows a transition for other 60 Hz displays, if other passive-entertainment-only (computer incompatible) 60 Hz formats under consideration are accepted.

Resolution scalability can be achieved though using a separate MPEG-2 picture stream for the resolution enhancement layer. Resolution scalability can take advantage of the B frame approach to provide temporal scalability in both the base resolution and enhancement layers.

The technique described here achieves many highly desirable features. It has been claimed by some of those who are involved in the U.S. advanced television process that neither resolution nor temporal scalability can be achieved at high definition resolutions within the 18.5 mbits/second available in terrestrial broadcast However, the technique described here achieves *both* temporal and spatial-resolution scalability within this available data rate.

It has also been claimed that 2Mpixels at high frame rates cannot be achieved without the use of interlace within the 18.5 mbit/second data rate available. This also appears to be incorrect. Further, the system described here achieves not only spatial and temporal scalability, but it can also provide 2Mpixels at 72 frames per second.

In addition to providing these capabilities, the system proposed here is also very robust. It is substantially more robust than the current proposal for advanced television. This is made possible by the allocation of most or all of the bits to the base layer when very stressful image material is encountered. Such stressful material is by its nature both noiselike and very rapidly changing. In these circumstances, the eye cannot see detail associated with the enhancement layer of resolution. Since the bits are applied to the base layer, the reproduced frames are substantially more accurate than the currently proposed advanced television system which uses a single constant higher resolution.

Thus, the system described here optimizes both perceptual and coding efficiency, while providing maximum visual impact. This system provides a very clean image at a resolution and frame rate performance that had been considered by many to be impossible. It is believed that the system described here is likely to outperform the advanced television formats being proposed by ACATS to the FCC. In addition to this anticipated superior performance, this system also provides the highly valuable features of temporal and resolution layering.